arXiv:0812.1072v4 [physics.soc-ph] 15 Jul 2009

Multiresolution community detection for megascale networks by information-based replica correlations

Peter Ronhovde and Zohar Nussinov
Washington University in St. Louis, Campus Box 1105, 1 Brookings Drive, St. Louis, Missouri 63130, USA
(Dated: July 15, 2009)

We use a Potts model community detection algorithm to accurately and quantitatively evaluate the hierarchical or multiresolution structure of a graph. Our multiresolution algorithm calculates correlations among multiple copies ("replicas") of the same graph over a range of resolutions. Significant multiresolution structures are identified by strongly correlated replicas. The average normalized mutual information, the variation of information, and other measures in principle give a quantitative estimate of the "best" resolutions and indicate the relative strength of the structures in the graph. Because the method is based on information comparisons, it can in principle be used with any community detection model that can examine multiple resolutions. Our approach may be extended to other optimization problems. As a local measure, our Potts model avoids the "resolution limit" that affects other popular models. With this model, our community detection algorithm has an accuracy that ranks among the best of currently available methods. Using it, we can examine graphs over 40 million nodes and more than one billion edges. We further report that the multiresolution variant of our algorithm can solve systems of at least 200 000 nodes and 10 million edges on a single processor with exceptionally high accuracy. For typical cases, we find a super-linear scaling, $O(L^{1.3})$ for community detection and $O(L^{1.3} \log N)$ for the multiresolution algorithm where L is the number of edges and N is the number of nodes in the system.

PACS numbers: 89.75.Fb, 64.60.Cn, 89.65.–s

I. INTRODUCTION

One focus in the study of complex networks is identifying suspected internal structure, and one characterization of such structure is in terms of "community" divisions within a model graph. A recent introduction to the "physics of networks" can be found in [1]. One feature of organized structure within these systems is that the community divisions can depend on the scale at which the system is examined. Different scales correspond to distinct community divisions at different internal community edge densities. For many systems, including those with hierarchical organization, a "multiresolution" approach [2] is needed to capture the overall structure and the relationships between the elements at different resolutions. Examples of such systems can include biological processes [3, 4], food webs [5], air transportation networks [4], and communication networks [6]. Thus, multiresolution methods are an important extension of problems in community detection.

Some measures and methods regarding community detection are reviewed in [7, 8]. Quality functions include modularity defined by Newman and Girvan [9], a Potts model originally proposed by Reichardt and Bornholdt (RB) [10, 11], our Potts model [12] that eliminates the random partition applied by RB, an application of a Potts model utilizing a mean-field approximation with "belief propagation" [13], and another measure "fitness" [14]. Other approaches include clique percolation [15, 16], spectral [17], continuous mapping to a conic optimization problem [18], "label propagation" [19, 20], dynamical [21, 22], and maximum likelihood [23]. Karrer *et al.* [24] defined a measure of robustness of community structure based on random perturbations. Some efforts enhance or expand applications to more general systems such as weighted networks [11, 12, 25], heterogeneous systems [12, 26], bipartite graphs [27, 28], overlapping nodes [10, 14, 15, 28, 29], and multiresolution methods.

The multiresolution algorithm presented in this paper (**1**) determines and *quantitatively* evaluates the relative strength of multiresolution structure(s) within a graph by examining the correlations among several independent solutions ("replicas") of the same graph over a range of resolutions. Strong correlations in the normalized mutual information (NMI) or the variation of information (VI) indicate the "best" system resolutions, and the relative value of the measure gives a quantitative estimate of the strength of the structures. This quantitative evaluation of the best resolution(s) for the system is lacking or missing in most other multiscale community detection algorithms. (**2**) The method is not limited to hierarchical structures but applies to general structures at different scales. (**3**) Our approach is based on relative information comparisons, so it can in principle be used with any community detection model that can target different resolutions. (**4**) The underlying Potts model and community detection algorithm demonstrate an accuracy at least equal to the best methods currently available (see Appendix A) [12]. The model is robust to the effects of noise (see Appendices A and B), and as a local measure, it is free of the "resolution limit" [30] as discussed in the literature [9, 11, 31, 32]. (**5**) With improvements discussed in Sec. IV, it is competitive with the best algorithms currently available both in terms of speed and possible system size. A single community solution can achieve sys-

tems as large as 40 million nodes and one billion edges with a computational time of 3.7 hours (see Appendix C) [33]. (**6**) Our multiresolution algorithm is extremely accurate for large systems (see Sec. VII). (**7**) We apply it to *megascale* systems with over 10 million edges and 200 000 nodes with a run time of about 4.6 hours on a single processor [33]. The algorithm should adapt very efficiently to parallel or distributed computing methods enabling larger systems to be studied.

Hierarchical organization is the most obvious type of multiresolution structure. Some earlier work on hierarchies in graphs can be found in [3, 34]. Examples of more recent efforts in analyzing hierarchical structures in graphs are [4, 6, 14, 35, 36]. Arenas *et al.* [36] defined a multiresolution method using modularity that makes novel use of the resolution limit [30]. Reichardt and Bornholdt [10], Arenas *et al.* [36], Kumpula and coworkers [37], and Heimo *et al.* [38] also study multiresolution applications of an RB Potts model.

In this paper we will show, for the first time, how information theory based measures may be used to systematically extract the best community partitions on all scales. This will enable us to methodically determine the hierarchical or multiresolution structure of arbitrary networks. In Sec. II, we first briefly review the information measures that we employ. Then in Secs. III and IV, we briefly discuss our Potts model and community detection algorithm, followed by an explanation of their applications to multiresolution analysis in Sec. V. We then present several examples in Sec. VI. The exceptional accuracy of the multiresolution algorithm is addressed in Sec. VII, and we conclude in Secs. VIII and IX. Details concerning the high accuracy and large size limit of the underlying community detection algorithm are relegated to Appendixes A and C respectively. Appendix B demonstrates an example of new transition effects in community detection (such transitions directly affect replica correlations). Appendix D explains a generalization of our replica method for other, nongraph theoretical, optimization problems. Appendix E elaborates on some details related to the benchmark accuracy test discussed in Sec. VII.

II. INFORMATION THEORY MEASURES

The normalized mutual information I_N and the variation of information V provide methods of comparing one proposed community division to another. In order to define $I_N(A,B)$ or $V(A,B)$ between two partitions A and B, we first ascribe a Shannon entropy $H(A)$ for an arbitrary community partition A. We assign the probability that a given node will fall in community k as $P(k) = n_k/N$, where n_k is the number of nodes in community k and N is the total number of nodes in the system. Then the Shannon entropy is

$$H(A) = -\sum_{i=1}^{q_A} \frac{n_k}{N} \log \frac{n_k}{N} \tag{1}$$

where q_A is the number of communities in partition A.

Mutual information $I(A,B)$ was developed within information theory. It evaluates how similar two data sets are in terms of information contained in both sets of data. The mutual information between two partitions A and B of a graph is calculated by defining a "confusion matrix" for the two community partitions. The confusion matrix specifies how many nodes n_{ij} of community i of partition A are in community j of partition B. Mutual information $I(A,B)$ is defined as

$$I(A,B) = \sum_{i=1}^{q_A} \sum_{j=1}^{q_B} \frac{n_{ij}}{N} \log \left(\frac{n_{ij} N}{n_i n_j} \right) \tag{2}$$

where n_i is the number of nodes in community i of partition A and n_j is the number of nodes in community j of partition B. An interesting generalized mutual information is also defined in [39]. Danon *et al.* [40] suggested that a normalized variant [41] of mutual information be adapted for use in evaluating similar community partitions. Using Eqs. (1) and (2), the normalized mutual information $I_N(A,B)$ between partitions A and B is

$$I_N(A,B) = \frac{2I(A,B)}{H(A) + H(B)} \tag{3}$$

which can take values in the range $0 \leq I_N(A,B) \leq 1$. Fred and Jain [41] introduced, for computer vision problems, a single resolution application of NMI that we use in our work.

The variation of information [42] is a metric in the formal sense of the term and measures the "distance" in information between two partitions A and B. Using Eqs. (1) and (2), $V(A,B)$ is calculated by

$$V(A,B) = H(A) + H(B) - 2I(A,B). \tag{4}$$

As an information distance, low values of $V(A,B)$ indicate better agreement between partitions A and B. VI has a range $0 \leq V(A,B) \leq \log N$. It is sufficient and even preferable to use the un-normalized version of VI. We utilize both NMI and VI to demonstrate that our approach is not limited to a specific measure.

The mutual information I and Shannon entropy H also play a supplemental role in determining multiresolution structure. For the Shannon entropy H, we average over all replicas using

$$\langle H \rangle = \frac{1}{r} \sum_{A} H(A). \tag{5}$$

For I_N, V, and I, we calculate the average of the measures over all pairs of replicas with

$$\langle S \rangle = \frac{2}{r(r-1)} \sum_{A>B} S(A,B) \tag{6}$$

where S is any of the information measures and r is the number of replicas. We use base 2 logarithms in all information calculations.

Similarly, higher order cumulants of S can be computed with a (replica symmetrically weighted) probability distribution function that we set to be

$$P(S) = \frac{2}{r(r-1)} \sum_{A>B} \delta\big[S - S(A,B)\big]. \qquad (7)$$

In Eq. (7), $\delta\big[S - S(A,B)\big]$ is the Dirac delta function. For any function f of S, the expectation value of f is

$$\langle f \rangle = \int dS\ P(S) f(S). \qquad (8)$$

Formally, in our probability distribution of Eq. (7), the information measure S plays a role analogous to the overlap parameter in spin-glass problems.

III. POTTS MODEL HAMILTONIAN

We briefly review our Potts model approach to community detection [12]. Generally speaking, community detection algorithms based on quality functions begin a community evaluation by measuring the number of connected nodes within (or outside) proposed communities. In general, these edges can be weighted or unweighted. The quality function must contrast this measure to some "expected" value or directly evaluate missing connections in some manner. If a linear addition of edge weights (connected and unconnected) is applied, the constructed model is equivalent to a Potts model spin system. As it applies to community detection, such a model was first proposed by RB [11], which demonstrated a clear bridge between community detection methods and statistical physics. RB implemented their model with a weighted comparison to a random partition (a "null" model) which included modularity as a special case.

In our Potts model, we directly sum the edge weights (connected and unconnected) in an energy calculation without a weighted null model. Thus we avoid a comparison to the properties of another graph, random or otherwise, and despite a global energy sum, we obtain an effectively local measure of community structure. As a local measure, the model is also free from the resolution limit discussed in the literature [30, 31, 32].

The general weighted Hamiltonian for our model is

$$\mathcal{H}(\{\sigma\}) = -\frac{1}{2} \sum_{i \neq j} \left(a_{ij} A_{ij} - \gamma b_{ij} J_{ij}\right) \delta(\sigma_i, \sigma_j) \qquad (9)$$

which we refer to as an "absolute Potts model" (APM). $A_{ij} = 1$ if nodes i and j are connected and are 0 otherwise. $J_{ij} \equiv (1 - A_{ij})$. The values $\{a_{ij}\}$ and $\{b_{ij}\}$ are general positive weights of the connected and unconnected edges, respectively, which allow both symmetric and directed graphs. $\{A_{ij}\}$, $\{J_{ij}\}$, $\{a_{ij}\}$, and $\{b_{ij}\}$ are all fixed by the definition of the system. γ is an externally defined weighting parameter for the unconnected edge weights. In practice, we use a symmetric matrix with integer weights (faster integer computations) on both connected and unconnected edges (γ is a rational number). σ_i is a Potts spin variable that can take an integer value $1 \leq \sigma_i \leq q$. The value of σ_i for a given node is the model equivalent of community membership. That is, node i is a member of community k if $\sigma_i = k$. The number of spin states q can be specified as a constraint or can be determined by the lowest energy configuration over all values of q. The Kroneker delta $\delta(\sigma_i, \sigma_j) = 1$ if $\sigma_i = \sigma_j$ and $\delta(\sigma_i, \sigma_j) = 0$ for $\sigma_i \neq \sigma_j$. As in [21, 43], the interaction between spins is attractive if they are connected and repulsive if they are not connected. A further important feature of the Hamiltonian is that each spin interacts only with other spins in the same community. The optimal ground state of Eq. (9) is often difficult to locate in practice, so we identify the communities of a system by searching for low-energy states of this Hamiltonian.

The edge density of a particular community k is $p_k = 2l/[n(n-1)]$ where l is the number of edges and n is the number of nodes in the community. We can relate the model weight parameter γ to the *minimum* internal edge density p_{in} for every community. We obtain this relation from a simple calculation on the minimum number of interior edges that results in an energy of zero or less for a single community. An alternative method is to calculate the minimum number of edges that will merge two connected communities. Then we can apply an inductive argument to establish the same inequality. For unweighted graphs, the relation is

$$p_{in} \geq \frac{\gamma}{\gamma + 1}. \qquad (10)$$

For a weighted graph, the relation is similar, $p_{in} \geq \gamma/(\gamma + \overline{w})$, where $\overline{w}$ is the average weight of connected edges within each community and p_{in} is then the edge weight density as compared to a maximally connected community with the same average weight $\overline{w}$. These density relations are useful because the typical internal community edge density is equivalent to the *resolution* of a system. As a result, the resolution for the graph as a whole is also effectively set by γ. This property is distinct from the resolution limit in the literature because the resolution set by this method is independent of a graph's own global parameters [30, 31, 32].

IV. COMMUNITY DETECTION ALGORITHM

We apply the Potts model of Eq. (9) with a simple community detection algorithm that is nevertheless extremely accurate, at least as accurate as the best available algorithms (see Appendix A) when used with our model [12]. The algorithm sequentially "picks up" each node and places it in the community that best lowers the energy based on the current state of the system. We repeat this process for all nodes and continue iterating until no moves are found after one full cycle through all

nodes. This part of the dynamical approach is similar to parts of algorithms used in [6, 19]. We can also choose to test the communities for possible merges that can arise due to local minima traps [44]. This test is more important for heavily weighted graphs with $\gamma \ll 1$. We can optionally further allow zero-energy moves for difficult problems. We attempt t independent optimization trials (generally $O(1)$) and select the lowest energy configuration as the solution. Each trial permutes the order in which the nodes are initially traversed. Appendix B illustrates the effect of additional trials using a common benchmark problem with increasing levels of noise.

The algorithm has been modified to use the intuitive neighbor-node search such as in [6, 19, 45] and a symmetric initial state of one node per cluster also in [6, 19, 45] and applied in a more general dynamical context in [21]. We further optimize the algorithm by allowing it to skip nodes that are already strongly defined within their respective communities. Empirically, we find that the neighbor search drastically improves performance for sparse graphs to $O(N^{1+\beta}Z^{1+\beta}t\log Z)$ for some small β [46, 47], where N is number of nodes and Z is the average node degree. The factor of $\log Z$ is due to a neighbor-node binary search for each connection matrix (A_{ij} or J_{ij}) evaluation. The factor of NZ is due to the iteration over all neighbors, and the factor of β in the exponent is due to the number of full NZ iterations which depends on the topology of the system, the initial state of the system, and the resolution being solved (*i.e.*, the model weight γ). This scaling enables us to achieve systems of at least $O(10^7)$ nodes and $O(10^9)$ edges for a single application of the algorithm. Details of one of the large tests are discussed in Appendix C. We have solved systems up to $O(10^5)$ nodes and $O(10^7)$ edges for the multiresolution algorithm [33] as discussed in Sec. VI C.

V. MULTIRESOLUTION ALGORITHM

One challenge in developing a multiresolution algorithm is that of selecting the best resolution(s) for the system. A straight-forward method that avoids the choice of resolution is to iteratively solve the system (with a necessary change in γ for our model) and collapse the communities into "supernodes" until the system is organized into a forced hierarchical structure. This approach is viable; but even when the system is hierarchical in nature, there is the question of whether the best resolutions were resolved at each stage. Our algorithm enables a quantitative analysis that determines the best resolutions and applies to general types of multiresolution structure.

A. Motivation

Ideally, we desire an algorithm that allows the system to communicate what the best resolutions are; but without *a priori* information, the correct weights for these resolutions are not obvious in general. In order to identify the proper resolutions, we examine information-based correlations among independent replicas (independent solutions) via NMI or VI over a range of resolutions. Rather than using the replicas to simply identify a unique optimized solution for each resolution, we examine correlations among the entire set. We then select the strongest correlations as the best resolutions.

From a global perspective, the average NMI (between all pairs of replicas) indicates how strongly a given structure dominates the energy landscape by measuring how well the replicas agree with each other. High values of the NMI (often manifested as peaks) correspond to more dominant, and thus more significant, structures. From a local perspective, at resolutions where the system has well-defined structure, a set of independent replicas should be highly correlated because the individual nodes have strongly preferred community memberships. Conversely, for resolutions "in-between" two strongly defined configurations, one might expect that independent replicas will be less correlated due to "mixing" between competing divisions of the graph. Random effects will usually reduce the correlations between independent solutions.

A similar argument applies to VI where, as an information distance, low values of VI correspond to better agreement among replicas. With these information-based correlations, we obtain a set of multiresolution partitions of the graph, but we also obtain an estimate of the relative strength of the structures at each resolution. Note that this argument does not distinguish between unrelated multiresolution structures or those that are strictly hierarchical in nature although nothing prevents the imposition of additional hierarchical constraints if desired.

Implicit in this argument is the idea that local minima in the energy landscape represent meaningful, even if perhaps incomplete, information about the graph. The same assertion was made in [4, 10] for modularity and the RB Potts model. Moderate levels of "confusion" caused by random or competing effects within a graph do not destroy information contained in the global energy landscape, and the replica correlations of our algorithm are a measure of the "complexity" of that landscape. As the noise in the system is increased we expect that the transition to incoherence (where replicas are weakly correlated) to occur rapidly (see end of Sec. VIII and a brief example of an accuracy transition in Appendix B). If an algorithm can verifiably solve for the global minima of a system in most cases, the problem of community detection is solved in principle. Since this is difficult to do in practice, the replica correlations in our algorithm take advantage of the fact that we cannot always locate the optimal ground state(s).

In principle, one can also include in Eq. (9) interactions between each of the r replicas to produce a "free energy" type functional of the form

$$F = \sum_i \mathcal{H}_i(\{\sigma\}) - T\sum_{i\neq j} S(i,j). \tag{11}$$

where $S(i,j)$ is an information-based measure (e.g., I_N, V, etc.) between all replica pairs and T is a scale for this information measure. $S(i,j)$ is maximized when the community partitions are identical in all replicas. This information theory measure formally plays a role analogous to entropy in a free energy functional. T then plays the role of a "temperature." Sans the first term, the minima of F in Eq. (11) produce highly correlated random configurations (a "random high temperature configuration" of the system which appears without change in all replicas). Our algorithm in this work will amount to initially minimizing the first term in F, *i.e.*, $\sum_i \mathcal{H}_i(\{\sigma\})$, for a set of fixed $\{\gamma_i\}$. Out of this set of replica configurations, we then ask for which γ_i do we find a maximum of the correlations, $\sum_{i\neq j} S(i,j)$, when this information theory measure is plotted as a function of γ. A more sophisticated version of our algorithm minimizes F directly with both terms included in each step. The information theory measures that we employ may also be written for other (non-graph theoretic) optimization problems with general Hamiltonians, or cost functions, $\mathcal{H}$ (see Appendix D).

B. Algorithm

We start the algorithm with a weighted or unweighted graph. In Eq. (10), p_{in} is the *minimum* internal edge density for each community, and it is equivalent to the *resolution* of the system when we minimize Eq. (9). The algorithm uses Eq. (9) to solve a range of resolutions $\{p_i\} = [p_0, p_f]$ (decrementing p_i) corresponding to a particular set of model weights $\{\gamma_i\} = [\gamma_0, \gamma_f]$ as determined by Eq. (10). It is almost always sufficient to have $\gamma_0 \lesssim 19$ since it corresponds to a *minimum* community edge density of $p_0 \geq 0.95$. The final weight γ_f is found when the system is completely reduced. A completely reduced system is one that is fully collapsed into one community or one where disjoint sub-graphs will not allow the system to collapse any further.

Each iteration, we decrement the density p_i by a small value $\Delta p = 0.05$ (or 0.025 for smaller graphs) and calculate the corresponding γ_i. After a threshold value (say $p_t = 0.1$), we scale p_i by a factor of $1/2$ (or $3/4$ for smaller graphs) in order to take sizable steps towards a fully reduced system (necessary for large systems). One could readily implement an adaptable step or "fill-in" process since the order of trials is irrelevant for the result.

The algorithm takes three input parameters: the number of independent replicas r that will be solved at each tested resolution, the number of trials per replica t, and the starting density which we set to be $p_0 \simeq 0.95$ corresponding to $\gamma_0 = 19$. The number of replicas is typically $8 \leq r \leq 12$ and is selected based upon how much averaging (over all replica pairs) is needed or desired. The number of trials t per replica is generally $2 \leq t \leq 20$. For each replica, we select the lowest energy solution among the t trials as was discussed in Sec. IV. The value of t is chosen based on how much optimization is necessary to identify a strong low-energy configuration [44].

The r replicas (and t optimization trials) are generated by reordering the "symmetric" initialized state of one node per community. That is, even though the initialized state is symmetric, the order that we traverse the list also affects the answer that we obtain. This occurs because the node-level dynamics of the underlying community detection algorithm in Sec. IV moves a node immediately upon identifying the best community membership given the current state of the system. Utilizing the r replicas, we then use the information-based measures of Sec. II to determine the multiresolution structure. Our algorithm is given by the following steps:

1. *Initialize the system.* Initialize adjacency matrices (A_{ij} and J_{ij}) and weights (a_{ij} and b_{ij}) based on the system definition. Use Eq. (10) and p_0 to calculate the initial model weight γ_0.

2. *Solve all replicas at this resolution p_i.* Initialize the current replica to a symmetric state of one node per community. Use Eq. (9) to solve each replica with model weight γ_i at a cost of $O(N^{1+\beta} Z^{1+\beta} t \log Z)$ per replica [44, 46]. Repeat the process independently for all r replicas. Each trial and replica randomly permutes the order in which nodes are initially traversed in the respective solutions.

3. *Calculate the replica I_N, V, I, and H information measures.* Use Eq. (1) to calculate H for all replicas and Eqs. (2)–(4) to calculate I, I_N, and V between all pairs of replicas for this resolution p_i [48]. Calculate the average (see Eqs. (5) and (6)) and the standard deviation for each measure.

4. *Decrement to the next resolution p_{i+1}.* If $p_i > 0.1$, decrement $p_{i+1} = p_i - 0.05$ or 0.025 for smaller graphs. If $p_i \leq 0.1$, $p_{i+1} = p_i/2$ or $3p_i/4$ for smaller graphs. Calculate the model weight γ_{i+1} by Eq. (10). Return to Step 2 until the system is not further reducible (fully collapsed or disjoint sub-graphs will not collapse).

5. *Evaluate results.* For the range of model weights $\{\gamma_i\}$, plot each average $I_{N,i}$, V_i, I_i, and H_i versus γ_i. Determine the strongest correlations (I_N high or V low) in these plots (see Figs. 2 – 4, 6, 8, and 10). These strongly correlated regions correspond to the best multiresolution structure(s) in the graph. If the correlation is less than "perfect" ($I_N < 1$ and $V > 0$), we choose the lowest energy replica to be the partition solution. One could also choose to construct a "consensus" partition between all of the replicas [19, 41] at each notable resolution.

We estimate that the number of resolutions $\{p_i\}$ required to adequately specify an arbitrary system scales as $O(\log N)$. The dominant scaling of the algorithm

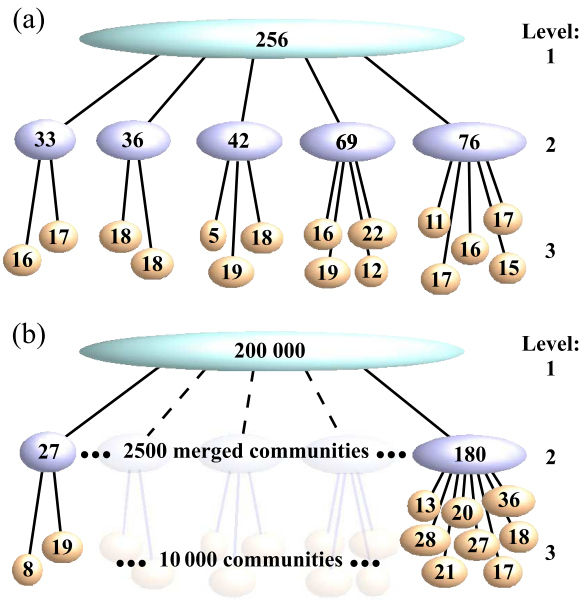

FIG. 1: (Color online) Heterogeneous hierarchical systems corresponding to the plots in Fig. 2 for panel (a) and the plots in Fig. 4 for panel (b). In panel (a), the 256 node system is divided into a three-level hierarchy where the unweighted edge connection probabilities at each level are the following: level 3 has $p_3 = 0.9$ between nodes in the *same* community with community sizes from 5 to 22 nodes (average 16). Level 2 has $p_2 = 0.3$ between nodes in *different* constituent sub-communities with merged community sizes from 33 to 76 nodes. Level 1 is the completely merged system of 256 nodes with $p_1 = 0.1$ between nodes in *different* sub-communities. The average edge density is $p = \overline{p}_1 = 0.182$. In panel (b), we increase the system size to 200 000 nodes. Level 3 has 10 000 communities with sizes from 6 to 37 nodes (average 20). Level 2 has 2500 communities with sizes from 27 to 180 nodes which are formed by merging two to eight communities from level 3. The density p_1 is changed from panel (a) to $p_1 = 0.000\,31$, and the average edge density is $p = \overline{p}_1 \simeq 0.0005$. This larger system has over ten million edges with approximately 62% of the edges being random noise between level 2 communities.

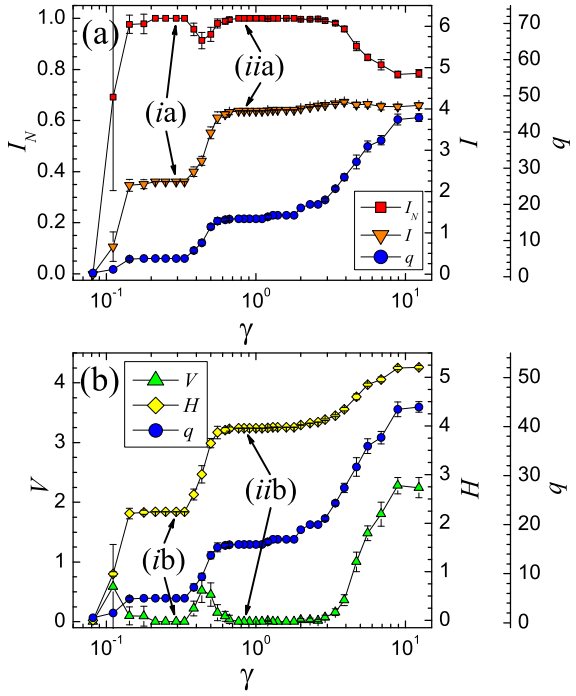

FIG. 2: (Color online) Plot of information measures I_N, V, H, and I in panels (a) and (b) vs. the Potts model weight γ in Eq. (9) for the three-level heterogeneous hierarchy depicted in Fig. 1(a). In panel (a), the squares represent the average replica normalized mutual information I_N (left axis), and the inverted triangles represent the average mutual information I (right axis). In panel (b), the triangles represent the average variation of information V (left axis), and the diamonds represent the average Shannon entropy H (right axis). For comparison, the circles in both panels (a) and (b) represent the average number of clusters q for the same set of replicas (right-offset axes). In panel (a) the peak I_N values (ia) and (iia) both accurately correspond to levels 2 and 3 respectively of the hierarchy depicted in Fig. 1(a). Similarly in panel (b) the minimum V values (ib) and (iib) also accurately correspond to levels 2 and 3, respectively, of the hierarchy. In panels (a) and (b), both the mutual information I and Shannon entropy H display a "plateau" behavior corresponding to the correct solutions. Plateaus in the average number of clusters q [51] also indicate important structures as in [36].

is almost always Step 2, so we estimate that the overall scaling is $O(N^{1+\beta}Z^{1+\beta}rt \log N \log Z)$ for some small β [46, 47].

Structures identified by this algorithm are not necessarily hierarchical; however, one can augment the algorithm by imposing an additional hierarchical constraint on some fraction of the replicas. Comparisons would then be made strictly between all pairs with and without this additional constraint. We applied this variation in both divisive and agglomerative approaches, but in our testing it only resulted in a modest improvement to the algorithm's ability to identify the best resolutions. Therefore, we use the above algorithm in order to take advantage of its generality and relative simplicity.

VI. EXAMPLES

We show the results of the multiresolution algorithm of Sec. V applied to several test cases [49]. In Secs. VI A and VI C, we illustrate a small 256 node and a larger 200 000 node hierarchy respectively with both systems depicted in Fig. 1. In Sec. VI B, we examine the structure of an Erdős-Rényi random graph for comparison to graphs with known internal structure. We then analyze two real social networks in Secs. VI D and VI E where the respective systems are depicted in Figs. 5 and 7. In Sec. VII, we also demonstrate the algorithm's exceptional accuracy for large systems.

A. 256 node hierarchy

The system in Fig. 1(a) depicts a set of 256 nodes for a constructed three-level heterogeneously-sized hierarchy. The results are seen in Fig. 2. The unweighted edge connection probabilities are p_k for $k = 1, 2, 3$. Level 3 has a density $p_3 = 0.9$ between nodes in the *same* community with community sizes from 5 to 22 (average 16) nodes. Level 2 has a density $p_2 = 0.3$ between nodes in *different* constituent sub-communities and is divided into five groups with merged sizes from 33 to 76 nodes. Level 1 is the completely merged system that has a density $p_1 = 0.1$ between nodes in *different* sub-communities. These edges provide some system noise. The average densities of communities at levels 1 and 2 are $p = \overline{p}_1 = 0.182$ and $\overline{p}_2 = 0.470$. We use eight replicas and four trials per replica at a total run time of 6.1 s [33].

In Fig. 2(a), the squares represent NMI averages over all replica pairs (left axis). The inverted triangles represent the mutual information I averages for the same replica pairs (right axis). In Fig. 2(b), the triangles represent VI averages over all replica pairs (left axis), and the diamonds represent the Shannon entropy H averages for the replicas (right axis). In both panels, the circles represent the average number of clusters across the replicas (right offset axes). All parameters are plotted versus the model weight γ where we use a logarithmic scale to facilitate comparing the behavior of a large range of system sizes from $N = 16$ nodes in Figs. 7 and 8 to as large as $N = 200\,000$ nodes in Figs. 1(b) and 4 [50].

The extrema (ia,b) and (iia,b) are the correctly determined levels 2 and 3 respectively of the test hierarchy depicted in Fig. 1(a). Peaks (ia) and (iia) have $I_N = 1$ and minima (ib) and (iib) have $V = 0$ which indicate perfect correlations among the replicas for both levels of the hierarchy. The "plateaus" in H and I are a second indication of the significant system structure whose importance will become more apparent in later examples. The plateau in the average q [51] is also an important indicator of system structure as used in [36]. However, Figs. 3, 6, and 8 discussed later demonstrate that some caution should be exercised when using the plateau criterion (in H, I, or q) for determining multiresolution structure.

At level 3 in Fig. 1(a), the average number of externally connected edges for each node is $Z_{out} \simeq 32.0$ with a random noise component of $Z_{out}^{noise} \simeq 19.8$. Both of these numbers are larger than the average number of internal edges, $Z_{in} \simeq 14.3$. Despite this imbalance, the algorithm easily identifies level 3 of the hierarchy because the external edges (particularly those due to the random noise) are not concentrated strongly enough into any one external cluster. This behavior is important for smaller communities on level 3 where Z_{out} is substantially larger than Z_{in}, and it illustrates that the model is robust to noise in the system.

The VI peaks at $\gamma_1 = 0.111$ and $\gamma_2 = 0.435$ in Fig. 2(b) correspond to the average inter-community edge densities, $p_1 = 0.1$ for sub-communities at level 2 and $p_2 = 0.3$ for sub-communities at level 3. Equation (10) relates the *minimum* internal edge density $p_{in} \geq \gamma/(\gamma + 1)$ for each community in a solved partition. We can arrive at this inequality, using inductive reasoning, by considering the minimum inter-community edge density required for two arbitrary communities A and B to merge. We apply the relation as an equality (*i.e.*, energy difference between the merged and unmerged states is approximately zero) for the peak VI values at γ_1 and γ_2. The respective densities are $p_1^{AB} = 0.100$ and $p_2^{AB} = 0.303$. These values correspond closely to the constructed inter-community densities p_1 and p_2 above. The local VI maxima show that "complexity" of the energy landscape increases at resolutions where $\gamma/(\gamma + 1)$ is equal to the mean inter-community edge density. The more intuitive interpretation is that the "complexity" of the energy landscape increases substantially when the energy difference between different states is approximately zero.

B. Erdős-Rényi random graph

In Fig. 3, for comparison purposes we show the results for a purely (Erdős-Rényi) random graph at the same average edge density $p = 0.182$ as the hierarchy in Figs. 1(a) and 2. We use eight replicas and four trials per replica at a total run time of about 6.9 sec [33]. The only peak (ia) in the random graph corresponds to a trivial division into groups with sizes of approximately $\{1, 2, 253\}$ among the various replica solutions. This peak indicates transitional behavior to lower density, essentially trivial, structures. Peaks such as (i) can be distinguished from more meaningful ones by the cluster size distribution or the corresponding information measures. The value of I at (ia) or V and H at (ib) all have very low information values. Otherwise, the random graph displays no significant multiresolution structure.

All of the information measures display a plateau behavior at (iia,b). The plateaus in NMI or VI do not indicate a clear multiresolution structure because the correlations are relatively poor ($I_N \simeq 0.70$ and $V \simeq 3.6$) for both measures. If we examine the detailed solutions across the plateaus (separate from our multiresolution algorithm), the average NMI and VI are $I_N = 0.644$ and $V = 4.04$ both of which indicate poor agreement. There is no consistent structure identified by the community detection algorithm in this region. Instead, the weak plateaus in NMI and VI indicate that the system is constrained within a set of similarly sized partitions that have similarly high community edge densities. This example also illustrates that if we use *only* the plateaus (in H, I, or q), there is a potential to incorrectly identify significant structure(s) in the system. This possibility can be remedied by information checks on nearby solutions in the plateau, but the poor NMI and VI correlations already appear to indicate the lack of consistent structure in the region.

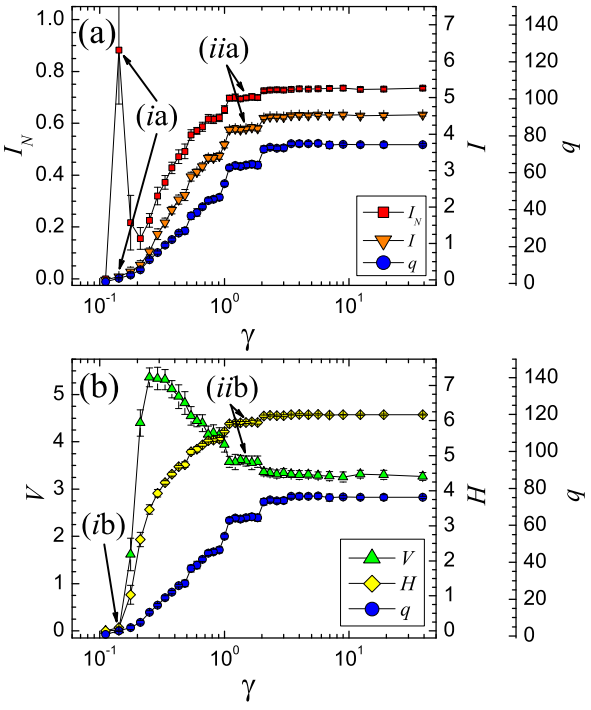

FIG. 3: (Color online) Plot of information measures I_N, V, H, and I in panels (a) and (b) vs the Potts model weight γ for a purely (Erdős-Rényi) random graph that has the same average density $p = 0.182$ as the hierarchy in Fig. 1(a) and the corresponding results in Fig. 2. The right-offset axes plot the number of clusters q. See Fig. 2 for a complete description of the legends and axes. In panel (a), the peak (ia) corresponds to a trivial partition of the system into groups with sizes of approximately $\{1, 2, 253\}$ among the different replicas. The trivial structure change in the NMI spike is indicated by its the low value of mutual information I at (ia) and by its low VI V and Shannon entropy H at (ib). The plateaus at (iia,b) do not correspond to a consistent multiresolution structure as evidenced by the poor NMI and VI correlations. Rather, they indicate multiple similarly sized configurations that have similar community edge densities.

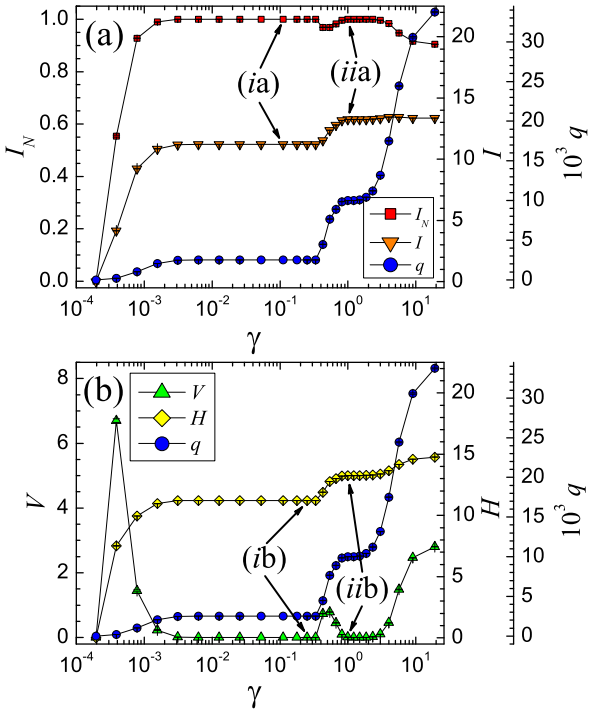

FIG. 4: (Color online) Plot of information measures I_N, V, H, and I in panels (a) and (b) vs. the Potts model weight γ for the large three-level heterogeneous hierarchy depicted in Fig. 1(b). The right-offset axes plot the number of clusters q. See Fig. 2 for a complete description of the legends and axes. With the exception of 15 weakly connected nodes (out of 200 000) and 5 merged clusters (out of 10 000) at (iia,b), the extremal values of I_N and V at (ia,b) and (iia,b) both accurately correspond to levels 2 and 3 respectively of the hierarchy depicted in Fig. 1(b).

C. Large hierarchy

A much larger hierarchy is depicted in Fig. 1(b). The system has 200 000 nodes and 10 011 428 edges. Approximately 62% of these edges are due to random noise between level 2 communities. For this system, $p_1 = 0.000\,31$, but $p_2 = 0.3$ and $p_3 = 0.9$ are unchanged from Fig. 1(a). There are 10 000 sub-communities at level 3 with sizes ranging from 6 to 37. Level 3 communities are combined in groups of two to eight to form the 2500 communities of level 2 with sizes ranging from 27 to 180. We use eight replicas and two trials per replica with a run time of about 4.6 hours [33]. In Fig. 4, extrema (ia,b) exactly identify level 2 of the hierarchy with perfect NMI and VI correlations, and extrema (iia,b) accurately identify ($I_N = 0.999\,995$ and $V = 1.42 \times 10^{-4}$) all but 5 merged clusters out of 10 000 and 15 nodes out of 200 000 nodes for level 3. Due to random fluctuations, all of these nodes have a random connectedness of 50% or less for their intended communities. This result is therefore consistent with the model and algorithm.

D. Dolphin social network

We tested a social network of 62 bottlenose dolphins in Doubtful Sound, New Zealand [52, 53, 54]. Three of the strongest partitions ((i), (iv), and (v)) are depicted in Fig. 5 using the results in Fig. 6. We use ten replicas with ten trials per replica at a total run time of about 0.78 sec [33]. We use a density scaling of 0.8 rather than 0.75 for $p_i < 0.1$ for Step 4 of the algorithm in order to more easily observe the transition between structures (i) and (ii) in Fig. 6. Configuration (i) identifies a grouping of 21 and 41 dolphins with perfect NMI and VI correlations ($I_N = 1$ and $V = 0$). This configuration agrees with an observed split of the dolphin network when a dolphin left the school [52], but our algorithm also suggests that this

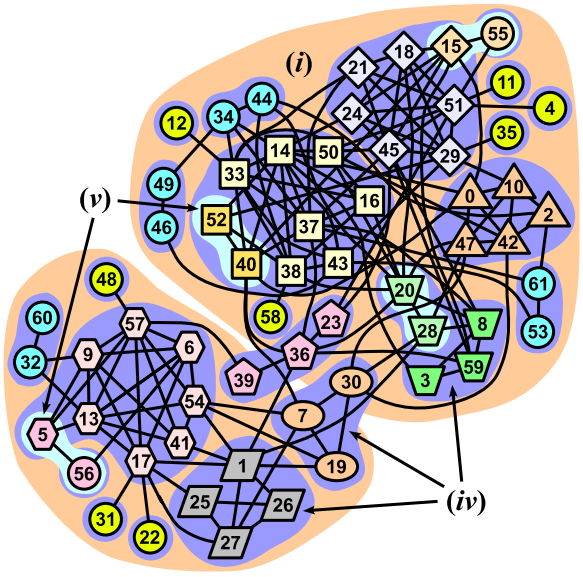

FIG. 5: (Color online) Pictorial representation of a social network of 62 bottlenose dolphins in Doubtful Sound, New Zealand [52, 53, 54]. These groupings correspond to structures (i), (iv), and (v) in Fig. 6 in order of smaller group sizes. The two-cluster partition (i) corresponds to a known split of the dolphin community [52]. In partition (iv), sub-groups are assigned distinct node shapes except for circles which indicate various one and two member groups. Structure (v) is identified from configuration (iv) when the four highlighted dyads of dolphins ($\{5, 56\}$, $\{15, 55\}$, $\{20, 28\}$, and $\{40, 52\}$) form distinct sub-groups. Note that sub-groups $\{7, 19, 30\}$ and $\{23, 36, 39\}$ in (iv) have nodes that are separated in their respective super-groups. These groups are examples of how our algorithm does not restrict node assignments between different resolutions, and they illustrate how the algorithm can apply to general types of multiresolution structure.

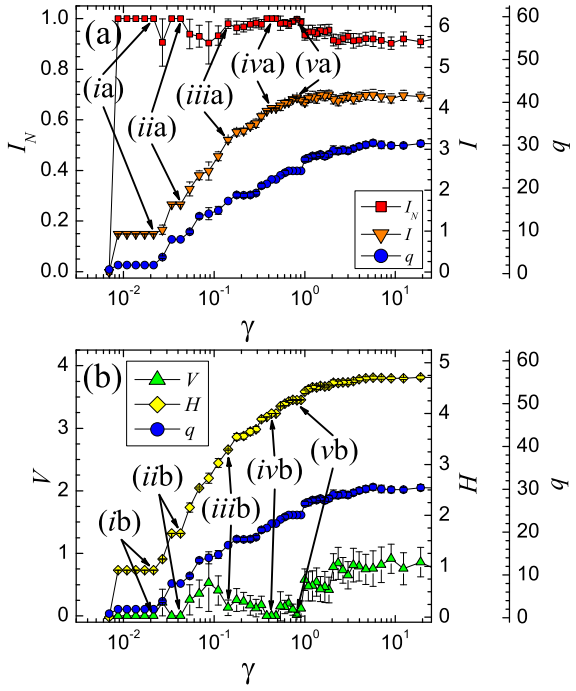

FIG. 6: (Color online) Plot of information measures I_N, V, H, and I in panels (a) and (b) vs. the Potts model weight γ for a social network of 62 bottlenose dolphins in Doubtful Sound, New Zealand [52, 53, 54]. A summary of results is depicted in Fig. 5 for configurations (i), (iv), and (v). The right-offset axes plot the number of clusters q. See Fig. 2 for a complete description of the legends and axes. One notable grouping is configuration (i) which corresponds to a known split of the dolphin community [52]. The structures represented by (ii) – (v) are other potential strongly defined partitions and are explained in the text.

configuration is not the only strongly defined partition for the system.

Our algorithm further identifies partitions (ii) – (v) as important candidate partitions based on the strong NMI and VI information correlations. Partition (ii) separates weakly connected dolphins ($\{4\}$, $\{11\}$, $\{12\}$, $\{35\}$, $\{58\}$, and $\{46, 59\}$) in the larger super-group of Fig. 5 into distinct sub-groups. Configuration (iii) is slightly less well-defined with information correlations of $I_N \simeq 0.980$ and $V \simeq 0.132$. It separates weakly connected dolphins ($\{22\}$, $\{31\}$, $\{39\}$, $\{48\}$, and $\{32, 60\}$) of the smaller super-group of partition (i) and also begins a coarse division of the larger super-group. Configuration (iv) is perfectly correlated and is the first major reconfiguration of both super-groups of structure (i). The data in the three largest groups of (iv) are largely divided along gender lines according to details presented in [53]. Configuration (v) is a slight variation of (iv) with $I_N \simeq 0.998$ and $V \simeq 0.0178$ which separates four dyads of dolphins ($\{15, 55\}$, $\{46, 49\}$, $\{32, 60\}$, and $\{20, 28\}$) into distinct groups. Among different tests, there is some variation in the predicted groupings where a few nodes can be reassigned between groups or separated into distinct communities. Sub-groups $\{7, 19, 30\}$ and $\{23, 36, 39\}$ of configuration (iv) have nodes that are split between the two super-groups of (i). These groups show that our algorithm does not restrict node assignments between different resolutions. This behavior allows our algorithm to solve general types of multiresolution structures.

All measures show a strong plateau for configuration $(i$a,b). The mutual information I shows weak plateaus at $(ii$a) and $(iv$a) but no plateau at $(iii$a) and $(v$a). Similarly, the Shannon entropy H shows weak plateaus at $(ii$b) and $(v$b) but no plateau for $(iii$b) and $(iv$b). The average number of clusters q as used in [36] also indicates the presence of structures (ii) and (v), but it misses partition (iv). Additionally, a weak plateau in q near configuration (iii) predicts a slightly different resolution than the extremal NMI and VI correlations. The weak plateau behavior of H, I, or q at different configurations of $(ii$a,b) – $(v$a,b) do not contradict the existence of valid structures. Rather, missing plateaus in the supplemental

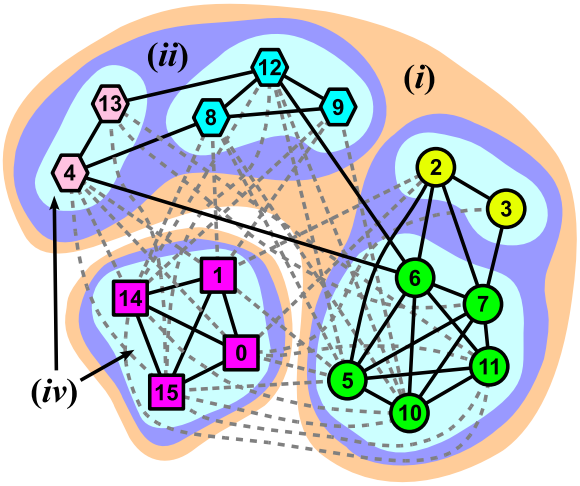

FIG. 7: (Color online) Pictorial representation of 16 Polopa tribes of Highland New Guinea [55, 56]. Solid lines represent allied relationships, and gray dashed lines represent antagonistic relationships. The three main levels of the structure are indicated by shaded areas. These groupings of tribes correspond to structures (i), (ii), and (iv) in Fig. 8 in order of smaller group sizes. Distinct node shapes (intermediate grouping) also correspond to structure (ii). The three-cluster structure (ii) corresponds exactly to the analysis in [55, 56]. Structure (iii) in Fig. 8 is formed when node 2 joins the group at the bottom-right of the figure.

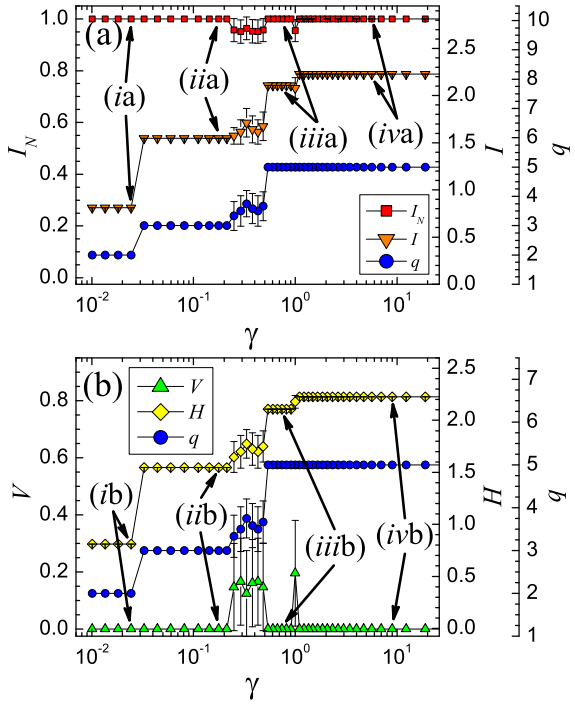

FIG. 8: (Color online) Plot of information measures I_N, V, H, and I in panels (a) and (b) vs. the Potts model weight γ for 16 Polopa tribes of Highland New Guinea. The results are summarized in Fig. 7. The right-offset axes plot the number of clusters q. See Fig. 2 for a complete description of the legends and axes. The most important structure represented in the figure is at $(ii$a,b) where the strong correlations agree exactly with data and analysis presented in [55, 56]. See the text for a full discussion of the other structures indicated in the figure.

measures H, I, or q can indicate a noisy graph in general or a strongly defined but transient resolution.

E. Highland Polopa tribe relations

Figures 7 and 8 show the results for 16 Polopa tribes of Highland New Guinea [55, 56]. These data feature allied, neutral, and antagonistic relations between the sub-tribes of the region. Hage and Harary [56] used symmetric edge weights of $+1$ for allied relations, 0 for neutral relations, and -1 for antagonistic relations in their analysis; but these "intuitive" weight assignments are inconsistent if extended to systems that include few or no antagonistic relations (such systems would tend to "collapse" into large groups). Therefore, our model uses the more consistent assignments of -1 for "neutral" relations and -2 for antagonistic relations. Interestingly, Hage and Harary [56] related the fact that the sub-tribes did not consider the possibility of strictly neutral relations among tribes. We use 12 replicas with 10 trials per replica to limit fluctuations in this very small data set at a total run time of about 0.46 sec [33]. We use an array data structure due to the missing edge weights.

Figure 7 depicts configurations (i), (ii), and (iv) from Fig. 8 in order of smaller group sizes. For presentation purposes, we allow three additional resolutions to be solved after the algorithm detects disjoint subgraphs at $(i$a,b). Our three-cluster partition (ii) agrees exactly with those discussed in [56]. All configurations indicated in Fig. 8 are strongly defined with $I_N = 1$ and $V = 0$. The first configuration (i) is a two-cluster solution which merges two sets of clusters of configuration (ii). The small size of the system causes the transition between configurations (i) and (ii) to be sharply defined. To resolve the ambiguity, we must reference the plateaus in the information measures H or I (or the number of clusters q [36]).

Strong NMI and VI values at $(iii$a,b) and $(iv$a,b) correspond to two five-cluster solutions. These solutions subdivide the three-cluster system into two slightly different dense configurations of allied tribes. In configuration (iii), node 2 is associated with the group on the bottom-right of Fig. 7. In configuration (iv), all groups are cliques (maximally connected sub-graphs). Both NMI and VI detect the transition between (iii) and (iv) with a short-lived spike. The information measures H and I also show the transition with plateaus at different values. Here, the number of clusters q does not detect the transition since q does not actually change. Again, this is due to the lim-

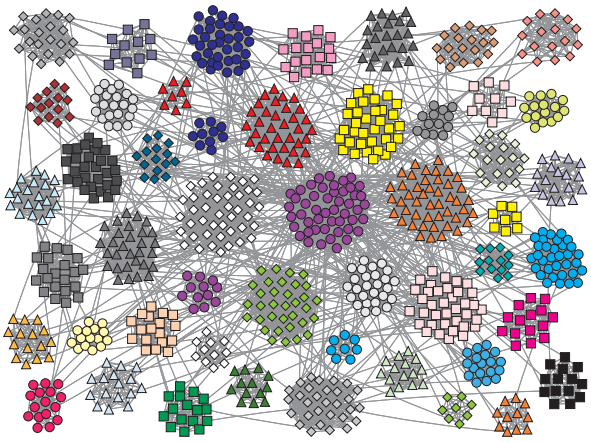

FIG. 9: (Color online) A sample graph with $N = 1000$ nodes from the new benchmark proposed in [57]. For presentation purposes, this depiction uses $\mu = 0.05$. Other parameters are $\alpha = 2$, $\beta = 1$, $\langle k \rangle = 15$, and $k_{max} = 50$ (see text).

ited variability in this system, but the same ambiguity occurs in Fig. 3 for all three supplemental measures H, I, and q.

VII. ACCURACY

In Figs. 9 – 11, we test the accuracy of the multiresolution algorithm of Sec. V with a recently proposed benchmark in [57]. An example graph with $N = 1000$ nodes is depicted in Fig. 9. This new benchmark can pose a significant challenge since it incorporates a more realistic heterogeneous distribution of community sizes and node degrees, and it allows for testing across a large range of system sizes. It divides a set of N nodes into q communities with sizes assigned according to a power-law distribution with an exponent β. The community sizes are optionally constrained by minimum and maximum sizes of n_{min} and n_{max}. The degrees of the nodes are also assigned in a power-law distribution with an exponent α with constraints specified by the maximum degree k_{max} and the mean degree $\langle k \rangle$. The minimum degree k_{min} is set so that the distribution gives the correct mean $\langle k \rangle$. A fraction $(1 - \mu)$ of the edges of each node are connected to nodes within their own communities. The remaining fraction μ are assigned to nodes in other communities.

We test systems with $N = 1000$ and 5000 nodes and power-law exponents of $\alpha = 2$ and 3 for the degree distribution and $\beta = 1$ and 2 for the community size distribution. We do not specify the optional community size constraints n_{min} or n_{max} allowing the benchmark program to specify them by the degree distribution. The node degree distribution is specified by $\langle k \rangle = 15$ and $k_{max} = 50$ where the mean degree $\langle k \rangle = 15$ was the most difficult of the tested values in [57]. We vary the mixing parameter μ in the range $0.1 \leq \mu \leq 0.7$. The accuracy

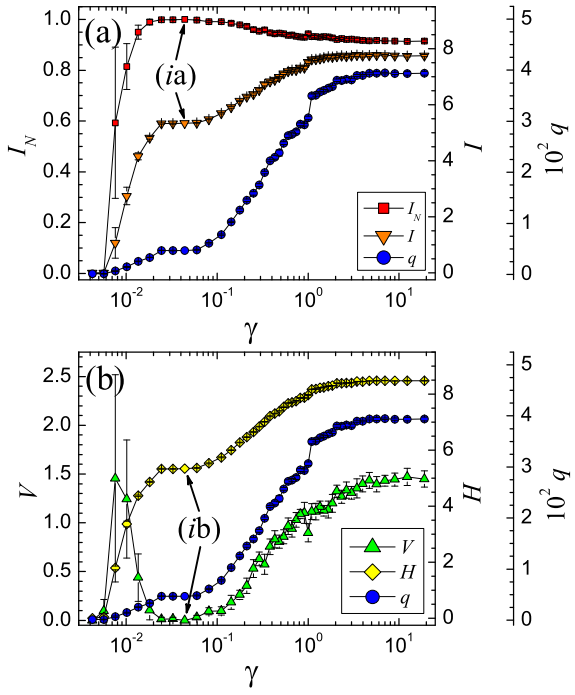

FIG. 10: (Color online) Plot of information measures I_N, V, H, and I in panels (a) and (b) vs. the Potts model weight γ for a single realization of the benchmark suggested in [57]. The right-offset axes plot the number of clusters q. See Fig. 2 for a complete description of the legends and axes. Figure 9 depicts a sample system from the benchmark (with a different mixing parameter μ) showing a distribution of community sizes. This example plot is for $N = 1000$ at $\mu = 0.5$ where 50% of each node's edges on average are connected to communities other than its own. We use $\alpha = 2$ and $\beta = 1$ for the power-law distribution exponents of the node degrees and the community sizes respectively. Using the algorithm in Sec. V, we identify the strongest NMI and VI replica correlations among the different resolutions as the "best" answer for the graph. For this graph at $\mu = 0.5$, there is only one extremal value of I_N and V which indicates that there is only one "best" resolution for the defined system (see also Appendix E). Note that these information values are the averages *among the replicas*. The full accuracy plot in Fig. 11 plots the average I_N between the "*best*" partitions and the *known* benchmark graphs for a range of the mixing parameter μ.

results are summarized in Fig. 11.

We apply the multiresolution algorithm of Sec. V to identify the "best" system partition. Figure 10 shows an application of the algorithm for a single benchmark graph with $N = 1000$, $\mu = 0.5$, $\alpha = 2$, and $\beta = 1$. In this plot, we identify the "best" system resolution by the strongest average NMI correlation between all pairs of replicas. We use $r = 8$ replicas with $t = 4$ energy optimization trials per replica. As seen in Fig. 10, both I_N and V (almost always) show only one extremal value which is the strongly defined system at (*i*a,b). Plateaus in H, I, and q qualitatively confirm the structure indicated by the extrema

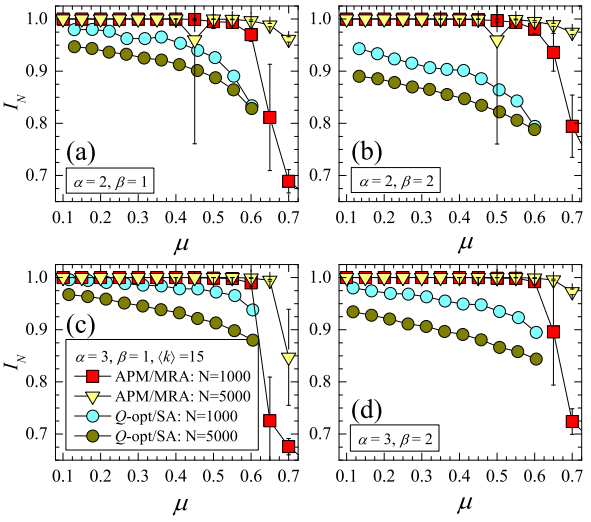

FIG. 11: (Color online) A plot of I_N vs. μ for a new benchmark problem proposed in [57]. I_N is calculated between the solved answer, by means the multiresolution algorithm in Sec. V using the absolute Potts model of Eq. (9), and the constructed benchmark graphs. An example multiresolution analysis for one generated graph is in Fig. 10. μ is the fraction of edges of each node (on average) that are assigned outside its own community. We tested the power-law distribution exponents $\alpha = 2$ and 3 and $\beta = 1$ and 2 for the node degrees and the community sizes, respectively. For comparison, we also plot the results from [57] determined by modularity optimization (Q-opt) using simulated annealing. With the APM, our multiresolution algorithm demonstrates extremely high accuracy for large systems (see text). Appendix E discusses the accuracy perturbations in panels (a) and (b) for $N = 5000$ nodes. Data for $N = 1000$ and $N = 5000$ nodes are averaged over 100 and 25 graphs respectively.

in I_N and V. From these data, we determine that there is only one "best" resolution for the defined system. See Appendix E for additional considerations in identifying the "best" benchmark resolution.

In Fig. 11, we identify the "best" partition for a set of benchmark graphs over a range of the mixing parameter $0.1 \le \mu \le 0.7$. We then compare each solution via NMI with the "known" partition. We average over 100 graphs for $N = 1000$ and over 25 graphs for $N = 5000$ for each tested μ. For comparison, we also include the results given in [57] for modularity optimization using a simulated annealing algorithm. Combined with the APM of Eq. (9), our multiresolution algorithm performs excellently, achieving almost perfect accuracy for each tested distribution exponent α and β and for a large range of the mixing parameter μ. The accuracy perturbations in panels (a) and (b) for $N = 5000$ nodes are due to benchmark graphs with more than one local extremum in I_N and V. These perturbations are a result of the automated selection of the single "best" resolution based on I_N and V extrema. We can largely eliminate them by a simple extension of the basic multiresolution algorithm (see Appendix E). They are also nearly eliminated for these values of N if we specify the default community size constraints of $n_{min} = 20$ and $n_{max} = 50$.

The absolute Potts model has little difficulty accurately solving the harder problem with $N = 5000$ nodes because the edges connected to external communities are spread over more communities on average. This construction causes a greater contrast of interior and external edge densities (considering edges connecting *pairs* of communities). This larger contrast allows the benchmark graph to be easily identified by the multiresolution algorithm. The converse occurs for small systems in the benchmark.

Our multiresolution algorithm has some difficulty in identifying all communities in this benchmark for exceptionally small systems ($N \lesssim 300$) where we achieve $I_N \simeq 1.0$ for a range of μ that increases with N (for $N = 300$, $I_N \simeq 1.0$ for $\mu \le 0.45$). Communities are partitioned locally, independent of any *global* parameters of the system; so this limitation is not a resolution limit effect. Rather, this behavior is due to simultaneously resolving communities with substantially different relative densities [58]. Palla *et al.* [15] stated that the community density should be used in identifying communities, which our Potts model does in effect. In Sec. III, we suggested that it is the typical community edge density that characterizes the *resolution* of a partition. The difficulty in this benchmark is due to defining communities by the fraction of each node's edges $(1 - \mu)$ that lie within its own community. Each community contains $\ell_s = n_s\langle k\rangle(1 - \mu)/2$ edges on average where n_s is the size of community s. The average edge density p_s of community s is

$$p_s = \frac{\langle k\rangle(1 - \mu)}{(n_s - 1)}. \tag{12}$$

The numerator is constant on average across all communities. Our Potts model solves heterogeneously-sized systems well (see Secs. VI A and VI C), but one notable implication of Eq. (12) is that the realistic distribution of community sizes leads to a substantial distribution of community edge densities with substantially different character for this benchmark.

Note also that our highly accurate results for $\mu = 0.6$ and 0.65 for most values of N, α, and β in Fig. 11 show that the concept of a weak community structure [59], where some nodes have more total edges connected to other communities than within their own, is not too restrictive because the external edges can be dispersed among many other communities. Indeed for $\mu > 0.5$, all clusters in this benchmark on average exceed the definition of a weak community since most, if not all, nodes have more exterior than internal edges. So-called weak communities can occur frequently in social networks for example. Individuals often know far more people than the size of the local "community" group(s) (friends, associates, etc.) of which they are members. We showed a similar, but more striking, result when identifying level 3 of the constructed hierarchy in Figs. 1(a) and 2 where the

smallest communities had many more external than internal edges. Nevertheless, the model could easily resolve the communities at the correct resolution.

VIII. DISCUSSION

In Figs. 2 – 4, 6, 8, and 10, strong correlations in NMI and VI appear to be consistent indicators of important multiresolution structures. In most cases the assessments of the "best" partitions are confirmed by "plateaus" in the mutual information I and the Shannon entropy H. These information plateaus are similar to those seen in the number of clusters q in [36] and that are also observed in our data [51]. In Ref. [36], the Arenas *et al.* indicated that plateaus in q correspond to the most relevant system structures. Our results largely affirm but also extend that observation.

In many pertinent applications of our algorithm, the final results (including, by fiat, our synthetic networks in Secs. VI A and VI C) are indeed hierarchical in the conventional sense. That is, solving the Hamiltonian of Eq. (9) anew with a different model weight γ may break the communities apart, but it does not swap vertices between different communities at the correct resolutions. As each resolution is solved independently in our algorithm, we may (and indeed do) find more complicated multiresolution partitions where node reassignments lead to overlaps between communities that are perhaps disjoint on another level. This latter case is more subtle and appears in systems such as the dolphin social network of Sec. VI D and other individually oriented networks.

Variations in run time scaling among the different tests is influenced, sometimes strongly, by different levels of effective noise in each system (aside from differing numbers of replicas and trials; see Appendix B). For example, the hierarchy for Fig. 2 had a run time of 6.1 s. The corresponding random graph in Fig. 3, with nearly the exact same density and number of nodes, finished in 6.9 sec.

NMI and VI possess different strengths for quantitatively assessing multiresolution structure. (**1**) Of course, NMI is normalized and VI is not (although one normalization for VI is $1/\log_2 N$ [42]). Both of these features are useful. (**2**) Figures 2–4 show that VI more clearly identifies poor configurations. In the high density regime ($\gamma \gtrsim 5$) of Figs. 2 and 4, NMI shows a lower correlation compared to the peak values at (i) and (ii); but VI clearly indicates poor agreement. In Fig. 3, VI in panel (b) visually indicates a much poorer correlation in the $\gamma \simeq 0.3$ region as compared to NMI in panel (a). (**3**) In Fig. 3(a), we identified peak (ia) as a "trivial" division with a huge component weakly connected to some small branch elements. If one was actually interested identifying these very low-density solutions, NMI does identify them. In panel (b), V and I simply indicate a very low-information configuration.

In many cases, extrema in either NMI or VI are sufficient to identify the multiresolution structure of a system. Occasionally, we need to additionally reference the mutual information I or the Shannon entropy H (or the number of clusters q [36]). For example, in Fig. 2 NMI and VI almost do not distinguish between the $\gamma = 0.83$ partition (the exactly correct one) and the $\gamma = 1.6$ partition (one weakly connected node separates to form a new community) because the separation between the two configurations is almost imperceptible. Both of these partitions correspond to level 3 of the hierarchy depicted in Fig. 1(a), and both partitions have perfect correlations ($I_N = 1$ and $V = 0$). In this case, the small changes in information measures H and I indicate a redundant $\gamma = 1.6$ partition. Also in Figs. 10 and 11, we used the plateau to distinguish, when needed, between strongly correlated transient partitions (due to random elements of the benchmark generation process) and the more stable partition corresponding to the intended solution.

A similar challenge can occur for very small systems, such as in the transition from (i) to (ii) in Fig. 8, or for systems with few intercommunity connections. As the resolution is adjusted in these systems, variability can be more limited; and system transitions can be sharply defined. For these systems, it is possible that the NMI and VI correlations can remain strong and constant while crossing a structural transition. In Fig. 8, we avoid this ambiguity by noting that H and I clearly show a transition between structures (i) and (ii). Such systems can also accentuate the perceived plateaus in the multiresolution data because the variation in different configurations is small and transitions between major configurations can be sharp.

Given the distinctions, the two evaluations of multiresolution structure ("plateau" behavior in H, I, and q or strongly defined I_N and V correlations) are complimentary. While the plateau behavior is important, it is a more qualitative assessment of the "best" resolutions for the system. At least for our Potts model, under some conditions the plateaus in H, I, or q can be weak enough to prevent them being used as the universal indicator of multiresolution structure. In Fig. 3, the plateaus even corresponded to a set of similarly sized partitions with similar densities rather than consistent structure. The NMI and VI approach can more easily identify short-lived, but nevertheless strongly defined, structures (such as configuration (iv) in Fig. 6) that the plateau criterion can miss. In all Figs. 2 – 4, 6, 8, and 10, the major benefit of using the NMI and VI evaluations is that it appears to give a *quantitative* estimate of the "best" resolutions. Together, the information measures appear to provide a consistent, accurate, and quantitative method of identifying general multiresolution structure.

In further work, we will also consider a different method of adjusting the resolution of the system using the Hamiltonian

$$\mathcal{H}_{vt}(\{\sigma\}) \;=\; -\frac{1}{2}\sum_{i\neq j}\left[\,(a_{ij}+\alpha_{ij})\,A_{ij} - (b_{ij}+\beta_{ij})\,J_{ij}\right] \times \delta(\sigma_i,\sigma_j) \tag{13}$$

where α_{ij} and β_{ij} are the new model weights as compared to γ in Eq. (9). This *variable topology* Potts Hamiltonian is a generalized and continuous version of threshold cut-offs in weighted graphs. It presents an alternative method of continuously scaling the system by using an additive rather than a multiplicative scaling. It differs from Eq. (9) in that it progressively adjusts the topology of the system where multiplicative scaling does not change the system's connectedness. Additive scaling may provide a different perspective on the evolution of the system structure over different scales, and it may better simulate how some real world models are "stressed."

Additionally, it may be possible to probe the system at a local level by using either localized partitions or by analyzing details within the confusion matrix at each resolution. With this approach, we may be able to identify stable, but localized, structures beyond the information conveyed in the global information-based correlations.

We discovered and will report in detail in an upcoming publication on a new sharp crossover between typical-easy and rare-hard community detection problems [46]. Our finding of a *community detection transition* constitutes an analog of the singular transition, or more precisely, a singular region in the k-SAT (satisfiability) problem. Mézard *et al.* [60] found that the hardest problems occur along well-defined loci in the phase diagram of random satisfiability problems. These loci of hard problems separate the SAT region (of satisfiable random problems) and the overly constrained UNSAT region (in which the constraints cannot all be simultaneously satisfied). We ascertained a similar phenomenon within community detection. See Appendix B for a summary of one facet of this transition.

Qualitatively, the analog of the SAT region is a common "easy" and "fast" community detection region. A "transition" region, where computational cost rapidly increases and accuracy rapidly decreases, corresponds to the singular region of the k-SAT problem. A "hard" and "slow" community detection region corresponds to the UNSAT region of the k-SAT problem. For some community detection problems, the convergence rate can accelerate in the hard region due to the problem being rapidly trapped by local energy minima.

In a future work, we will detail the minimization of the "free energy" type functional of Eq. (11). This functional contains both the Potts model energy and the composite information function. This latter information theory measure is maximized when the correlation between replicas is maximal.

IX. CONCLUSION

We use a Potts model measure for community detection and apply it to detecting multiresolution structures: (**1**) Our approach identifies and *quantitatively* evaluates the 'best' multiresolution structure(s), or lack thereof, in a graph. (**2**) All resolutions are solved independently, so the algorithm allows for the identification of completely general types of multiresolution structure. (**3**) It is based on information comparisons, so in principle is should apply to any community detection model that can examine different resolutions. (**4**) The underlying Potts model and algorithm are as accurate as the best methods currently available (see Appendix A). The model is a local measure of community structure, so it is free from the 'resolution limit' as discussed in the literature [12, 30, 31, 32, 36, 37]. (**5**) Building on this foundation, the multiresolution algorithm demonstrates extremely high accuracy for large systems using a recent benchmark proposed in [57] (see Sec. VII). (**6**) We estimate that the computational cost scales as $O(N^{1+\beta}Z^{1+\beta}rt\log N\log Z)$ for some small β [46, 47] where r is the number of replicas, t is the number of optimization trials per replica, Z is the average node degree, and N is the number of nodes. We have tested our community detection algorithm on systems as large as $O(10^7)$ nodes and $O(10^9)$ edges (see Appendix C) [33]. The multiresolution algorithm requires a substantial number of individual community solutions; but due to the speed of the underlying algorithm, it can nevertheless examine systems over $O(10^5)$ nodes and $O(10^7)$ edges on a single-user workstation. The algorithm should extend very efficiently to parallel or distributed computing methods allowing larger systems to be studied.

ACKNOWLEDGEMENTS

We thank UCINet and M. E. J. Newman for network data made available on their websites. This work was supported by the LDRD DR on the physics of algorithms at LANL.

APPENDIX A: ACCURACY OF THE COMMUNITY DETECTION ALGORITHM

We demonstrate the accuracy of the community detection algorithm in Sec. IV that is used to calculate the individual replica solutions in Step 2 of the multiresolution algorithm discussed in Sec. V. Our results using this frequent model problem in the literature were previously presented in [12]. The constructed model has 128 nodes divided into 4 clusters with 32 nodes each. For each node, Z_{in} edges are randomly connected to other nodes within its own community and Z_{out} edges are randomly connected to nodes in one of the other three communities. The total degree of each node is $Z = Z_{in} + Z_{out}$ where we require an average degree of $Z = 16$.

The task is to verify the defined community structure. In Fig. 12, we use $\gamma = 1$ in Eq. (9) with q constrained to four. We plot the percentage of correctly identified nodes p versus the average number of externally connected edges per node Z_{out}. We use the same measure of the "percentage" of correctly placed nodes as Ref. [19]

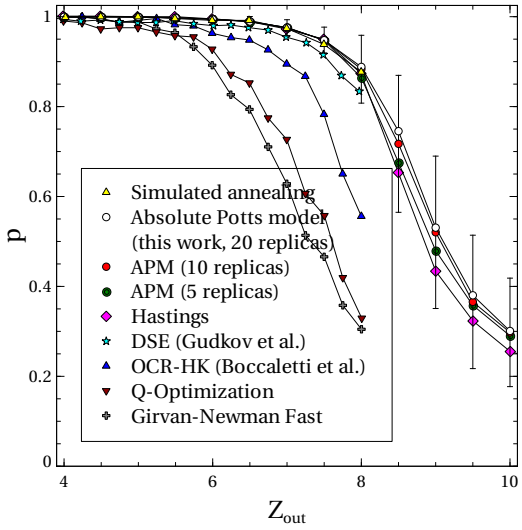

FIG. 12: (Color online) Reproduced from Ref. [12]. A plot of the percentage of correctly identified nodes p versus Z_{out}, the average number of edges that each node has connected to nodes outside of its own community. The average number of total edges per node is $Z = 16$. The benefit of extra trials t reaches a point of diminishing returns around $t = 10$ for many tests, and it is the intermediate difficulty trials ($8 \leq Z_{out} \leq 9$) that benefit the most from the additional optimization trials. Note that the accuracy of our APM of Eq. (9) and algorithm in Sec. IV is at least equal to the best algorithms. Each point is averaged over 500 systems.

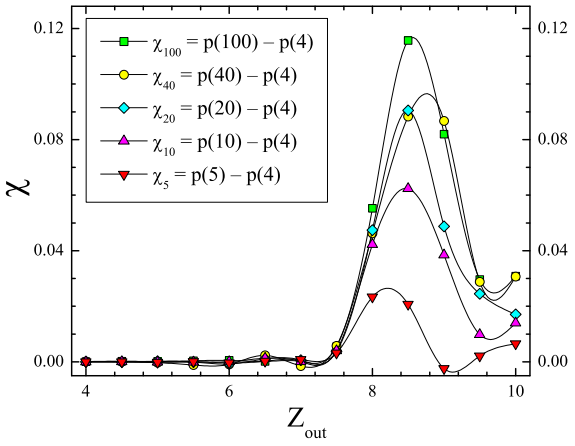

FIG. 13: (Color online) A plot of the susceptibility $\chi_n \equiv p(t = n) - p(t = 4)$ versus Z_{out}, the average number of edges that each node has connected exterior to its own community. χ_n is the percentage increase in the accuracy of each test as the number of trials $t = n$ is increased from $n = 5$ to $n = 100$. The average number of total edges per node is $Z = 16$. p is the percentage of correctly identified nodes from Fig. 12. The curves are spline fits and are intended for visualization purposes only. Additional trials are unnecessary in the easy region $Z_{out} \lesssim 7$. The benefit of extra trials is largest in the short transition region $8 \leq Z_{out} \leq 9$. Afterwards, the benefit diminishes into the hard region $Z_{out} \gtrsim 9.5$ where the accuracy improvement is small even with a large number of attempted optimization trials.

within [61]. Four sets of data in Fig. 12 were assimilated by Boccaletti *et al.* [22]. Simulated annealing [40] proved to be the most accurate algorithm in [22] although it is computationally expensive. Hastings [13] and Gudkov *et al.* [21] also demonstrated accurate results.

Our results in Fig. 12 use an older, slower version (without the neighbor-node search) of our algorithm. For a small system of only 128 nodes and $q = 4$ by constraint, the difference in run time would be small. For many of the tests, the benefit of extra trials t reaches a point of diminishing returns by $t = 10$. High noise systems rapidly trap different replicas in local energy minima, so it is the "intermediate" difficulty solutions ($8 \leq Z_{out} \leq 9$) that benefit the most from additional optimization trials. Our method maintains an accuracy rate at least equal to the best available algorithms. In particular, it maintains a 95% or better accuracy rate up to $Z_{out} = 7.5$.

APPENDIX B: TRANSITION EFFECTS OF NOISE LEVEL ON COMMUNITY DETECTION ACCURACY

The benchmark problem that serves as the basis for data in Fig. 13 is discussed in detail in Appendix A. In Fig. 13, we plot for several numbers of trials n, the "susceptibility" $\chi_n \equiv p(t = n) - p(t = 4)$ versus Z_{out}, the average number of edges that each node has connected exterior to its own community. The average number of total edges per node is $Z = 16$. p is the percentage of correctly identified nodes from Fig. 12 (see Ref. [19] in [61]), and t is the number of trials at each test. The ordinate χ in Fig. 13 is the percentage improvement in accuracy based on the number of optimization trials that are used.

As Z_{out} increases, the noise in the system increases. Figure 13 illustrates how the noise in the system affects the effort required to solve the system as accurately as possible. The benefit of extra optimization trials is negligible for the easy region up until about $Z = 7$. Additional trials become more important for a short transition region ($8 \leq Z_{out} \leq 9$). Afterwards, the benefit of additional trials quickly reaches a point of diminishing returns in the hard region $Z_{out} \gtrsim 9.5$ where it fails to produce large improvements in accuracy despite significantly more computational effort.

As the number of trials n increases, the "susceptibility" χ_n progressively exhibits a more pronounced peak. Such a trend is also evidenced in the susceptibility of finite size physical systems. We have also identified a similar and related dynamic feature of the transition that is quantified by the increased computational time required for a single solution [46] (beyond any added computational cost due to extra energy optimization trials).

APPENDIX C: COMMUNITY DETECTION OF A LARGE SYSTEM

We tested our community detection algorithm in Sec. IV using the neighbor-node search on a synthetic network with over one billion links. We generated a random set of $N = 40$ million nodes separated into 1.25 million heterogeneously-sized communities with sizes ranging from 10 to 62 nodes. (Note that it is the number of edges that limits the calculation as opposed strictly to the number of nodes.) The random edge connection probability for the communities was $p_{in} = 0.9$. Nodes between these communities were connected with a probability of $p_{out} = 5.31 \times 10^{-7}$. Each node has an average number of interior and exterior edges of $Z_{in} \simeq 29$ and $Z_{out} \simeq 21$. The total number of edges was 1 000 211 862. The average density of the graph was $p = 1.25 \times 10^{-6}$.

We used $\gamma = 1$ in Eq. (9) and the algorithm in Sec. IV to solve the system. There were 13 nodes that were not placed within their intended communities. These are likely due to random initialization fluctuations. The information correlations for the "known" and solved answers were $I_N \simeq 1.00$ and $V = 1.85 \times 10^{-6}$ with $V_{max} = \log_2 N \simeq 25.3$. Both of these measures indicate very strong agreement. The total calculation time was 3.7 h on a single processor [33].

APPENDIX D: GENERALIZATION OF THE INFORMATION-BASED REPLICA METHOD

In Sec. V, we may recast the information theory measures used to evaluate the correlation between different replicas for other (non-graph theoretic) optimization problems with general Hamiltonians (or cost functions) $\mathcal{H}$. An alternate form of Eq. (2) for the mutual information between replicas i and j is

$$I(i,j) = H(i) + H(j) - H(i,j) \tag{D1}$$

where $H(i)$, $H(j)$, and $H(i,j)$ denote the entropies of replica i, replica j, and the combined system formed by both replicas, respectively. Instead of using Eq. (2), we write the Shannon entropy $H(i,j)$ for the combined replicas i and j which we then apply in Eq. (D1). For general Hamiltonians $\mathcal{H}$, we replace $H(i)$, $H(j)$, and $H(i,j)$ by a thermodynamic entropy for the respective systems.

In the general case, the thermodynamic entropy $H(i,j)$ of the system formed by the union of replicas i and j is

$$H(i,j) = \frac{\partial}{\partial T}\left\{\beta^{-1}\log\left[\mathrm{Tr}_{i,j}\left(e^{-\beta\mathcal{H}(i)} + e^{-\beta\mathcal{H}(j)}\right)\right]\right\}, \tag{D2}$$

and the entropy $H(i)$ of system i or j is

$$H(i) = \frac{\partial}{\partial T}\left\{\beta^{-1}\log\left[\mathrm{Tr}_{i}\left(e^{-\beta\mathcal{H}(i)}\right)\right]\right\}. \tag{D3}$$

$\mathcal{H}(i)$ and $\mathcal{H}(j)$ are the Hamiltonians of replicas i and j, and $\beta = 1/(T\ln 2)$ is the inverse temperature. Within our approach, an ensemble reduces to a finite number of points (replicas) whose correlations are monitored by information theory measures. This form pertains to the general case in which both i and j pertain to a collection of decoupled copies, and the traces are over all coordinates in replicas i and j.

The standard mutual information of Eq. (2) is generally not invariant (as it ideally should be) under the permutation of "identical" nodes (those with an identical neighbor list that are otherwise indistinguishable by other parameters of the system). Specifically, we refer to nodes i and j as identical in a graph if the adjacency matrix A is invariant under the permutation of node i with node j [62]. That is, A commutes with the permutation of nodes i and j, $[P_{ij}, A] = 0$, if nodes i and j are identical. The thermodynamic entropies of Eqs. (D2) and (D3) are invariant under permutations of identical nodes because any symmetries, or lack thereof, are fully represented in the system Hamiltonian $\mathcal{H}$.

In the simplest case with only one copy of the system in replica i and one copy in replica j, there is only one term in both i and j; and the designation $\mathrm{Tr}_{i,j}$ becomes redundant (the entropies of i and j are also trivially $H(i) = H(j) = 0$). In a more realistic approximation to thermodynamic quantities, each of the replicas i and j contain a number of independent decoupled copies of the system. Inserting Eqs. (D2) and (D3) into Eq. (D1), we obtain the mutual information between i and j. NMI and VI are then given by Eqs. (3) and (4). Other information measures $S(i,j)$ between replicas i and j may also be computed. Along similar lines, multi-replica (higher than two) forms may replace the sum over two-replica configurations in Eqs. (11) and (D2).

We may also reconstruct the information measures using a different physical analogy. The Shannon entropy of Eq. (1) is analogous to an ensemble where each of the N nodes corresponds to one point in the ensemble. The communities correspond to q possible states of a single particle with energies $\{E_k\}$ for $k = 1$ to q at a given temperature T such that the same community occupation probabilities are reproduced as $p_k = n_k/N = e^{-\beta E_k}/\sum_{i=1}^{q} e^{-\beta E_i}$ where the inverse temperature is $\beta = 1/(T\ln 2)$. The mutual information I of Eq. (2) is equivalent to an ensemble of size N for a two-particle system in which each particle can be in any of q states. The interaction between the two particles is such that it leads to energies $\{E_{ij}\}$ for the two occupied communities i and j. These interactions lead to a relative probability $p_{ij} = n_{ij}/N$ for occupying the two-particle states that is proportional to $e^{-\beta E_{ij}}$. The effective Hamiltonian for the resulting physical system does not directly depend on the identities of the N nodes (although it does not distinguish between "identical" and distinguishable nodes).

One potential limitation of our thermodynamic framework in Eqs. (D2) and (D3) is that general, non-graph theoretic, applications may require many copies of the same system. The traces Tr_i, Tr_j need to be calculated

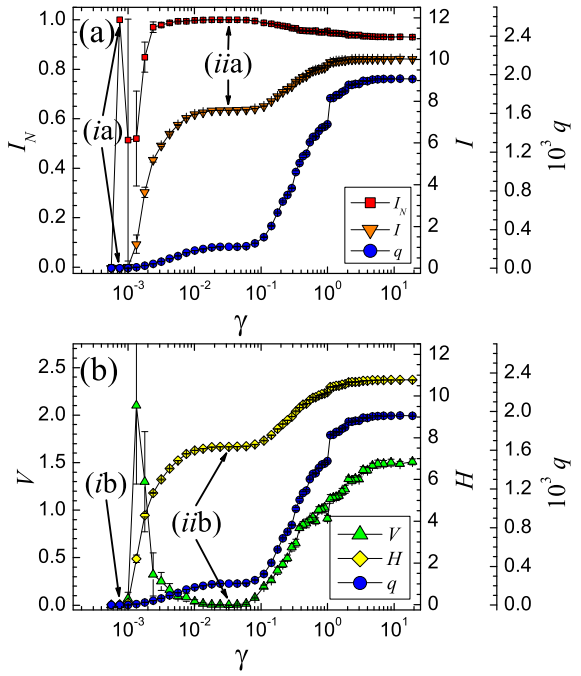

FIG. 14: (Color online) Plot of information measures I_N, V, H, and I in panels (a) and (b) vs the Potts model weight γ for a single realization of the benchmark suggested in [57]. The right-offset axes plot the number of clusters q. See Fig. 2 for a complete description of the legends and axes and Sec. VII for a full explanation of the benchmark. This example plot is a multiresolution analysis for $N = 5000$ at $\mu = 0.45$ where 55% of each node's edges on average are connected to communities other than its own. The power-law distribution exponents for the node degrees and the community sizes are $\alpha = 2$ and $\beta = 1$, respectively. We use the algorithm in Sec. V to attempt to identify the "best" resolution for the graph. For some cases in the benchmark, such as this graph at $\mu = 0.45$, there exists more than one extremal value of I_N and V where the low-density configuration at (ia,b) also has slightly stronger NMI and VI correlations ($\delta I_N \simeq 6.3 \times 10^{-5}$) than the intended benchmark answer at (iia,b). In this example case, a casual inspection indicates that the stable region at (iia,b) is clearly the "best" partition which also corresponds almost exactly to the intended benchmark solution. The automated version of the algorithm favors the slightly stronger low-density configuration at (ia,b) as the "best" resolution for the graph.

on multiple copies of the same system. This is bypassed in the application of mutual information for graph problems because the node number N effectively plays the role of many ensemble points (multiple replica copies) on which the thermodynamic average is to be taken.

APPENDIX E: MULTIRESOLUTION BENCHMARK COMMENTS

As discussed in Sec. VII, we used the new benchmark problem presented in [57] to test the accuracy of our multiresolution algorithm of Sec. V. Our algorithm attempts to identify all strongly defined resolutions. By design, the benchmark in [57] constructs a "realistic" system with a single intended solution; however, random effects of the graph generation process can also create additional transient, but nevertheless strongly defined resolutions which our algorithm can detect. In implementing the benchmark, we endeavor to automate the identification process to determine the single "best" resolution as intended by the benchmark. We explain two special cases.

The first difficulty is encountered for $\mu \lesssim 0.4$. We can detect multiple resolutions with perfect correlations ($I_N = 1$ and $V = 0$) for resolutions near the intended benchmark solution which occur more frequently as μ decreases. This effect is similar to partition (i) that occurred near partition (ii) in Fig. 8. The transitional resolutions are not necessarily meaningless partitions on an individual graph-by-graph basis, but they are artifacts of the randomly generated system and thus vary across the different benchmark graphs. Similar to structure (iia,b) in Fig. 14, the plateaus in the information measures H or I (or the number of clusters q [36]) indicate a more "stable" partition. It is this stable partition that corresponds to the intended solution for the benchmark graph. Thus, when necessary, we use the plateaus to discriminate between the short-lived and the most stable strongly defined partitions in order to determine the single "best" resolution for each benchmark graph.

A second difficulty is shown in Fig. 14 which occurs most frequently in the range of mixing parameter $0.45 \lesssim \mu \lesssim 0.65$. The stable configuration that corresponds to the intended benchmark answer is configuration (iia,b). The low-density, transient, but strongly correlated configuration at (ia,b) has a slightly higher NMI correlation. Even a casual visual inspection of the data in Fig. 14 indicates that configuration (iia,b) is the dominant configuration for the graph. Specifically, configuration (iia,b) possesses both very strong NMI and VI correlations ($I_N \simeq 1.0$ and $V \simeq 0.0$) as well as stable and long H, I, and q plateaus, and indeed it corresponds almost exactly to the intended benchmark answer. However, the automated application of the multiresolution algorithm slightly favors configuration (ia,b) as the "best" resolution since it has a higher NMI ($\delta I_N \simeq 6.3 \times 10^{-5}$) and a lower VI. (See Secs. VI B and VIII regarding potential problems of using the plateaus in H, I, or q as the primary measure for identifying the "best" resolutions.)

These graphs are the cause of the accuracy perturbations in Figs. 11(a) and 11(b). They are less frequent for $\beta = 2$ since the community size distribution is more skewed towards smaller communities than for $\beta = 1$. We note that the average accuracy for the perturbations in Figs. 11(a) and 11(b) is still high at $I_N \simeq 0.96$. In

Fig. 11, an iteration cap acted as an effective filter for most low-density spikes. We could further improve the automated analyses of such graphs by replacing this filter with moving NMI or VI averages (*i.e.*, each moving average is over the NMI or VI of several nearby resolutions) to exclude resolutions such as the short-lived configuration (i).

[1] M. E. J. Newman, Phys. Today **61(11)**, 33 (2008).

[2] We adopt the term "multiresolution" as used in [37] to indicate that this algorithm is not limited to hierarchical structures.

[3] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási, Science **297**, 1551 (2002).

[4] M. Sales-Pardo, R. Guimerà, A. A. Moreira, and L. A. N. Amaral, Proc. Natl. Acad. Sci. U.S.A. **104**, 15224 (2007).

[5] A. Clauset, C. Moore, and M. E. J. Newman, Nature (London) **453**, 98 (2008).

[6] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, J. Stat. Mech.: Theory Exp. (2008) P10008.

[7] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).

[8] S. Fortunato and C. Castellano, in *Encyclopedia of Complexity and Systems Science,* edited by R. A. Meyers (Springer, New York, 2009), e-print arXiv:0712.2716.

[9] M. E. J. Newman and M. Girvan, Phys. Rev. E **69**, 026113 (2004).

[10] J. Reichardt and S. Bornholdt, Phys. Rev. Lett. **93**, 218701 (2004).

[11] J. Reichardt and S. Bornholdt, Phys. Rev. E **74**, 016110 (2006).

[12] P. Ronhovde and Z. Nussinov, e-print arXiv:0803.2548.

[13] M. B. Hastings, Phys. Rev. E **74**, 035102(R) (2006).

[14] A. Lancichinetti, S. Fortunato, and J. Kertész, New J. Phys. **11**, 033015 (2009).

[15] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, Nature **435**, 814 (2005).

[16] J. M. Kumpula, M. Kivelä, K. Kaski, and J. Saramäki, Phys. Rev. E **78**, 026109 (2008).

[17] M. E. J. Newman, Phys. Rev. E **74**, 036104 (2006).

[18] R. Hildebrand, e-print arXiv:0806.1896.

[19] U. N. Raghavan, R. Albert, and S. Kumara, Phys. Rev. E **76**, 036106 (2007).

[20] M. J. Barber and J. W. Clark, e-print arXiv:0903.3138.

[21] V. Gudkov, V. Montealegre, S. Nussinov, and Z. Nussinov, Phys. Rev. E **78**, 016113 (2008).

[22] S. Boccaletti, M. Ivanchenko, V. Latora, A. Pluchino, and A. Rapisarda, Phys. Rev. E **75**, 045102(R) (2007).

[23] A. Clauset, M. E. J. Newman, and C. Moore, in *Proceedings of the 23rd International Conference on Machine Learning* (Association of Computing Machinery, New York, 2006).

[24] B. Karrer, E. Levina, and M. E. J. Newman, Phys. Rev. E **77**, 046119 (2008).

[25] M. E. J. Newman, Phys. Rev. E **70**, 056131 (2004).

[26] L. Danon, A. Díaz-Guilera, and A. Arenas, J. Stat. Mech.: Theory Exp. (2006) P11010.

[27] R. Guimerà, M. Sales-Pardo, and L. A. N. Amaral, Phys. Rev. E **76**, 036102 (2007).

[28] N. Du, B. Wu, B. Wang, and Y. Wang, in *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology* (IOS Press, Amsterdam, 2008), Vol. 1, p. 176.

[29] Overlapping nodes are those that are shared between multiple communities.

[30] The resolution limit causes a community division of affected methods to be roughly constrained by the graph's own global parameters. For modularity, the number of communities q scales as $\sqrt{L}$ on average [31] where L is the number of edges in the graph. This behavior is distinct from the utility of varying *graph-independent* model weights to examine system structure at different community edge densities. Arenas *et al.* [36] recently applied this resolution "limit" in a novel manner to examine the structure of a system over a range of resolutions, but the model used here is independent of this additional global constraint.

[31] S. Fortunato and M. Barthélemy, Proc. Natl. Acad. Sci. U.S.A. **104**, 36 (2007).

[32] J. M. Kumpula, J. Saramäki, K. Kaski, and J. Kertész, Eur. Phys. J. B **56**, 41 (2007).

[33] Graphs were solved on a single core of an Intel Core 2 Quad Q6600 at 2.4 GHz with 2 GB random access memory (RAM) or on Opteron systems at 2.2 – 2.8 GHz with 4 – 32 GB of RAM.

[34] E. Ravasz and A.-L. Barabási, Phys. Rev. E **67**, 026112 (2003).

[35] L. da F. Costa and R. F. S. Andrade, New J. Phys. **9**, 311 (2007).

[36] A. Arenas, A. Fernández, and S. Gómez, New J. Phys. **10**, 053039 (2008).

[37] J. M. Kumpula, J. Saramäki, K. Kaski, and J. Kertész, Fluct. Noise Lett. **7**, L209 (2007).

[38] T. Heimo, J. M. Kumpula, K. Kaski, and J. Saramäki, J. Stat. Mech.: Theory Exp. (2008) P08007.

[39] V. Gudkov and V. Montealegre, Physica A **387**, 2620 (2008); *Complexity, Metastability, and Nonextensivity*, AIP Conf. Proc. No. 965 (AIP, New York, 2007), p. 336; V. Gudkov and S. Nussinov, e-print arXiv:cond-mat/0209112.

[40] L. Danon, A. Díaz-Guilera, J. Duch, and A. Arenas, J. Stat. Mech.: Theory Exp. (2005) P09008.

[41] A. L. N. Fred and A. K. Jain, in Proceedings of the IEEE Computer Society Conference on Computer Vision Pattern Recognition (Computer Society, Toronto, 2003), Vol. 2, p.128.

[42] M. Meilă, J. Multivariate Anal. **98**, 873 (2007).

[43] V. Gudkov, S. Nussinov, and Z. Nussinov, e-print arXiv:cond.mat/0209419.

[44] Additional work during each replica solution (Step 2 of the multiresolution algorithm) includes two optimizations. First, as in [12] each replica can utilize multiple trials t to mitigate the effects of local energy minima by selecting the lowest energy trial for that replica. Typically $t \simeq 4$, but it can be higher $t \simeq 20$ for difficult systems. Second, the algorithm manually merges clusters that will lower the energy of the system. This effect arises because the algorithm evolves the system by sequentially moving

one node at a time to a new community based on the largest energy decrease for the node given the current state of the system. Some edge configurations (particularly heavily weighted graphs with $\gamma \ll 1$) can hinder the process of merging clusters when moving one node at a time. The number of merges is indefinite, but it is generally small enough to not significantly alter the overall performance of the algorithm. In some cases, additional node-level adjustments may follow as a result of any manual merges. The computational cost is estimated to be $O(NZ \log Z \log q)$ except for $\gamma \ll 1$, where q is small.

[45] A. Clauset, M. E. J. Newman, and C. Moore, Phys. Rev. E **70**, 066111 (2004).

[46] The scaling dependence of β is a complicated function $\beta \equiv \beta(N, q, Z_{in}, Z_{out}, \gamma)$, where N is the number of nodes, q is the number of communities, γ is the Potts model weight, and Z_{in} and Z_{out} are the average number of interior and exterior connected edges for each node with respect to its own community. A typical average scaling is $\beta \lesssim 0.3$. Empirically for large graphs, worst case behavior is seen in localized regions with an intermediate to high Z_{out}/Z_{in} ratio where there is significant system confusion (where any algorithm would have difficulty). In these cases $\beta \simeq 9$, and the system is in transition to a structure that is more difficult to solve. The transition effect is localized and is not representative of the global scaling. Afterward, β drops to a value more representative of the average scaling. As the system becomes increasingly difficult, bordering on incoherent, the convergence rate can actually accelerate as the algorithm is rapidly trapped by local minima. The rapid transition is representative of general transitions between typical-easy and rare-hard problems and constitutes an analog of the singular transition point the k-SAT problem. We will report on this effect in an upcoming publication.

[47] We can eliminate the scale factor of $\log Z$ for any unweighted graph. However, here we wish to demonstrate the full weighted scaling, so almost all of our unweighted graph examples include the $\log Z$ scale factor.

[48] For the I_N, V, and I calculations, we are generally comparing configurations that are somewhat similar. The confusion matrix can have at most N non-zero entries; therefore, we use a pseudo-sparse matrix representation to calculate these measures usually in $O(N)$. Worst case (pathological) behavior for initialization is $O(Nq)$. The calculation cost can be as fast as $O(q)$ for strongly correlated systems. The total cost for Step 3 of the algorithm will typically scale as $O(Nr^2 \log N)$ where the user has control over the number of replicas r and $\log N$ is the estimated scaling for the required number of resolutions.

[49] Data for the constructed systems examined in this paper (Figs. 1 – 4) can be found at http://physics.wustl.edu/zohar/communitydetection/.

[50] We use a log scale for the model weight γ in part because γ is related to the minimum community edge density p_{in} given by Eq. (10). As a result, γ effectively specifies the *resolution* of the partition. A log scale better captures the variation in distinct configurations for $0 < \gamma \lesssim 19$ which covers the practical range of significant resolutions. In particular, we are interested in important low-density resolutions with $\gamma \ll 1$ *and* higher density resolutions with $1 \leq \gamma \leq 19$. The log scale does not change the extremal values of I_N and V that indicate the "best" resolutions; but compared to a linear scale in γ, it could affect the visual interpretations of the "plateaus" in the supplemental measures H, I, or q.

[51] We plot the *average* q as opposed to q based on an optimal partition as used in [36] in order to be consistent with our use of averaged information measures.

[52] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, Behav. Ecol. Sociobiol. **54**, 396 (2003).

[53] D. Lusseau, Proc. R. Soc. London, Ser. B **270**, S186 (2003).

[54] D. Lusseau, Evol. Ecol. **21**, 357 (2007).

[55] K. Read, Southwestern Journal of Anthropology **10**, 1 (1954).

[56] P. Hage and F. Harary, *Structural Models in Anthropology* (Cambridge University Press, Cambridge, 1983), pp. 40 and 56–60.

[57] A. Lancichinetti, S. Fortunato, and F. Radicchi, Phys. Rev. E **78**, 046110 (2008).

[58] Different internal community densities are best resolved at different values of γ (*i.e.*, different *resolutions*). As the number of nodes N decreases in the benchmark [57], the density of inter-community edges between two communities increases relative to the internal community densities. Together, they cause the range of the solvable resolutions (see the solution at (*i*a,b) in Fig. 10) to decrease and eventually restrict an accurate or intended solution.

[59] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, Proc. Natl. Acad. Sci. U.S.A. **101**, 2658 (2004).

[60] M. Mézard, G. Parisi, and R. Zecchina, Science **297**, 812 (2002).

[61] M. E. J. Newman, Phys. Rev. E **69**, 066133 (2004).

[62] As a specific realization, consider a symmetric adjacency matrix A for a four node graph given by $A_{12} = A_{13} = A_{14} = A_{23} = A_{24} = 1$. The permutation P_{34} that exchanges 3 with 4 leaves A invariant, so the graph itself is invariant. If we have two solutions B and C that place i and j in different communities then $I_N(B, C) \neq 1$ despite the invariance of A under the permutation. Unless one can distinguish between the nodes via external means, community groupings of the four nodes such as (123)(4) and (124)(3) correspond to the same breaking of the lattice (the graph is symmetric under the permutation), but they have a relative mutual information that differs from the self mutual information $[(8 - 3\log 3)/4$ vs $(10 - 6\log 3)/4$ for the two groupings respectively].